

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 26, 2001, 10:36:00 ; Search time 13.34 Seconds
(without alignments)
1286.292 Million cell updates/sec

Title: US-09-497-967-7
Perfect score: 2540
Sequence: 1 MKNILVILISFINQIKS.....QCDFANFLSILLISYLL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	236.5	9.3	667	1	TS1L_GIALA
2	228.5	9.0	713	1	TSAN_GIALA
3	189	7.4	328	1	CL70_GIALA
4	188	7.4	1609	1	LMG1_HUMAN
5	188	7.4	5376	1	ZAN_MOUSE
6	185.5	7.3	1696	1	PKC5_MOUSE
7	184.5	7.3	687	1	VS41_GIALA
8	183.5	7.2	1680	1	FUR2_DROME
9	179.5	7.1	1877	1	PKC5_MOUSE
10	179	7.0	2715	1	GL56_PARPR
11	178.5	7.0	1607	1	LMG1_MOUSE
12	175.5	6.9	3075	1	LMAL_HUMAN
13	174.5	6.9	3635	1	LMA5_MOUSE
14	173	6.8	3712	1	LMA_DROME
15	171	6.7	1895	1	YLK3_CAEEL
16	167.5	6.6	2704	1	GL68_PARPR
17	166.5	6.6	969	1	PAC4_HUMAN
18	166	6.5	3084	1	LMAL_MOUSE
19	165	6.5	1700	1	BAR3_CHITE
20	158	6.2	3106	1	LMA2_MOUSE
21	157	6.2	3110	1	LMA2_HUMAN
22	155.5	6.1	1246	1	YMW2_CAEEL
23	155.5	6.1	3672	1	LML2_CAEEL
24	153.5	6.0	1790	1	LMB1_DROME
25	150.5	5.9	670	1	VG50_HSV1
26	150.5	5.9	1169	1	YK82_YEAST
27	150	5.9	1168	1	LMB3_MOUSE
28	148.5	5.8	1639	1	LMG1_DROME
29	147	5.8	712	1	FBL1_CAEEL
30	146.5	5.8	937	1	PAC4_RAT
31	146.5	5.8	1193	1	LMG2_HUMAN
32	145.5	5.7	600	1	SP96_DICDI
33	145.5	5.7	790	1	ANP_NOTCO

34	142	5.6	1167	1	XMRK_XIPMA
35	141.5	5.6	610	1	LEM2_HUMAN
36	141.5	5.6	1786	1	LMB1_MOUSE
37	141	5.6	2911	1	FBN2_HUMAN
38	140.5	5.5	1416	1	YH81_CAEEL
39	140	5.5	1064	1	FBP1_STRPU
40	139.5	5.5	677	1	SP87_DICDI
41	139	5.5	768	1	LEM3_MOUSE
42	138	5.4	2437	1	NOTC_BRARE
43	138	5.4	2871	1	FBN1_MOUSE
44	136.5	5.4	1172	1	LMB3_HUMAN
45	136	5.4	2907	1	FBN2_MOUSE

ALIGNMENTS

RESULT 1

TS1L_GIALA

ID AC Q03185; STANDARD: PRT; 667 AA.

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE MAJOR SURFACE TROPHOZOITE ANTIGEN 11 PRECURSOR.

GN TSPL1.

OS Giardia lamblia (Giardia intestinalis).

OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.

OX NCBI_TaxID=5741;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ISOLATE AD-1;

RX MEDLINE=93241215; PubMed=8479449;

RA Ey P.L., Khanna K., Manning P.A., Mayrhofer G.;

RT "A gene encoding a 69-kilodalton major surface protein of Giardia intestinalis trophozoites";

RL Mol. Biochem. Parasitol. 58:247-258(1993).

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE PLASMA MEMBRANE.

CC -!- DOMAIN: CONTAINS 27 REPEATS OF A CXXC MOTIF.

CC -!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

CC -----

CC EMBL: M95814; AAA02687.1; -.

CC PIR: A48579; A48579.

CC HSSP: P02468; IKLO.

CC InterPro: IPR000561; EGF-like.

CC InterPro: IPR002174; Furin-like.

CC SMART: SM00181; EGF 3.

CC SMART: SM00261; FU; 5.

CC Antigen; Repeat; Transmembrane; Signal.

FT SIGNAL 1

FT CHAIN 18 667 MAJOR SURFACE TROPHOZOITE ANTIGEN 11.

FT DOMAIN 18 633 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 634 662 POTENTIAL.

FT DOMAIN 663 667 CYTOPLASMIC (POTENTIAL).

FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 667 AA; 68475 MW; 1DD9572703232B8D CRC64;

Query Match 9.3%; Score 236.5; DB 1; Length 667;

Best Local Similarity 22.7%; Pred. No. 6e-10;

Matches 128; Conservative 49; Mismatches 193; Indels 195; Gaps 30;

Oy 7 VILIISFINQIKSANCPCVGTETNTAGQVDDLTGTPANCVCQ-----KNEY 52

```

DB 1 MLLAIFYFVISTLAKTQTQTCBAECMEVGEITEITRCQTGKVPIDGKCVATANAN 60
QY 53 YNNA-----AFVPGASTCTP-----CP-QKDGAGQPNPAPATANLVQCNV 93
DB 61 CNASGDGDANQVCKMSSVPGNTLCTVSPDGVCSVAANEYFVPPNADATHDSVVSCE 120
QY 94 KCPAGPA-----TAGGAT-----DYAALITECVNCRINYNENAPNPNAGASTC 137
DB 121 ETPHLLANNKQVIGVAGCATCSAPKAPGSDNTPKAACTCKAAGFL--HTP--SEGLSSC 176
QY 138 -TACPNRVGGALTAGNAATIAQCNVACPTCTALDDGVTTDYVRSFTECVKCRNLNFYN 196
DB 177 EETCEPGYFGHTATAESKKT-----CK-STGSS-----EAPNVKIGDCLAC----MYN 222
QY 197 GNGNTPFPNPGKSCQTPCPAIPKPNVAQATLGNDAF--ITAQCNVA-----CPDG----- 244
DB 223 EASGNT-----LTCEKSAQKPSL-DKTSNDCCTGQNCATFCSSSGGDCGDCSGFILD 275
QY 245 -----TISAAGVNNVAQNTCTNCAPNFYNNAPNPNPGNSTCLPCPANKDYGA 294
DB 276 GQNCVKSDCXTENCKACTNPKAANEVCTECI-----STHHLTPTSQCVOYCOALGNYYA 329
QY 295 EATAGGAATLAKOCNTA-----CPDGTAIASGATNYVILQTECLNCAANFYFDGNFQAGS 350
DB 330 GTNADNKA-CKECTVANCKTND-----QGQCOTCNDGFKYKNGDAGSPCH 374
QY 351 SRCKACPANKVQAVATAGTATILIAQALECPAGIVL---TDGTTSTYKQ----- 398
DB 375 ESKTCSA-----GTA-----SDCT-ECPTGKALYKNGDGTGTCGCGTTCQGGSG 419
QY 399 -----AASCVKVC-----AANFYTTKOTDWMVAGIDTCTSC----- 428
DB 420 AKTCTGLTIDGASVCSGDQNEYPQNGICTSTARIVATCKNSNVANGI--CSSOTNGF 477
QY 429 -----NKKLTSGAEAN 439
DB 478 LRMNGGVETTKPKGKSVCEGANAD 502

RESULT 2
TS44_GIALA STANDARD; PRT; 713 AA.
AC P21849;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR SURFACE-LABELED TROPHOZOITE ANTIGEN 417 PRECURSOR.
GN TSA 417.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 30957 / WB;
RX MEDLINE=90280395; PubMed=2352929;
RA Gillin F.D., Hagblom P., Harwood J., Alek S.B., Reiner D.S.,
RA McCafferty M., So M., Guiney D.G.;
RT "Isolation and expression of the gene for a major surface protein of
RT Giardia lamblia."
RT Proc. Natl. Acad. Sci. U.S.A. 87:4463-4467(1990).
RN [2]
RP SEQUENCE OF 480-620 FROM N.A.
RC STRAIN=AD-1;
RX MEDLINE=93314970; PubMed=8325510;
RA Ey P.L., Mayrhofer G.;
RT "Two genes encoding homologous 70-kDa surface proteins are present
RT within individual trophozoites of the binucleate protozoan parasite
RT Giardia intestinalis."
RL Gene 129:257-262(1993).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
CC PLASMA MEMBRANE.
CC -!- DOMAIN: CONTAINS 29 REPEATS OF THE CXXC MOTIF.

```

```

CC -!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M33641; AAA02688.1; -.
CC EMBL; M97488; AAA02581.1; -.
CC PIR; A35502; A35502.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR002174; Furin-like.
CC SMART; SM00181; EGF; 1.
CC SMART; SM00001; EGF; like; 1.
CC SMART; SM00261; FU_3.
CC SMART; SM00001; EGF; like; 1.
CC Signal; Antigen; Glycoprotein; Transmembrane; Repeat.
CC FT SIGNAL 1 17
CC FT CHAIN 18 713
CC FT MAJOR SURFACE-LABELED TROPHOZOITE
CC FT ANTIGEN 417.
CC FT EXTRACELLULAR (POTENTIAL).
CC FT POTENTIAL.
CC FT CYTOPLASMIC (POTENTIAL).
CC FT N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD A -> T (IN STRAIN ADELAIDE-1).
CC FT VARIANT 582 582
CC FT VARIANT 606 606
CC FT SEQUENCE 713 AA; 72510 MW; 9AD7195843DE5601 CRC64;
CC SQ

Query Match 9.0%; Score 228.5; DB 1; Length 713;
Best Local Similarity 24.4%; Pred. No. 2.4e-09;
Matches 122; Conservative 34; Mismatches 170; Indels 175; Gaps 33;

QY 23 CPVGTETTAGOVDDLGTTPANCVCQKN-FYYN-----NAAAFVPGASTCTP-----CPQ 71
DB 69 CP-----QHSAGK-----CTCGGNSFMFKDGYSSGEGLPGLSLSSDGDGVCTE 115
QY 72 KKDAGAAPN---PPATANLVQCNVK---PAGTATAGGATDYAAIITECVNCRINFYEN 126
DB 116 -----AAPGYFAPVGAAN-TEQSVIACGDTTGTVAAGNTYKGIADCAEC-----S 161
QY 127 APNFNAGA-----STCTACPNRVGGALTAGNAATIVAOCN----- 162
DB 162 APDATAGAEAGVATCTKGVSK-----YLDNVVDKACQNSGNTKFFAVDDSENGKC 217
QY 163 VACPT-----GTALDDGVTTDYVRSFTECVKCRNLNFYNGNNTIP-----FNP-- 206
DB 218 VSCSDNLNGVANCDCSYDEQSKIKCTKTDNNYLTSTSEKVCVKDCKDGFPPKD 277
QY 207 ---GKSQCTPCPAKP--ANVAQATL--GNDATITACQNVACPDG----- 244
DB 278 DSSAGNKCLPCNDSTDIANCAATCALVSGRGAALVTC--ACTDGYKPSADKTTCEAVSN 336
QY 245 --TISAAGVNNVAQNTCTNCAPNFYNNAPNPNPGNSTCLPCPA--NKDYCAEATAG 300
DB 337 CKTPGCKACSKNEGKNEVCVDCDGSY-----LTP-TSQCIDSCAKIGNYYGA---TEG 386
QY 301 AATLAKOCNTA---CPDGTAIASGATNYVILQTECLNCAANFYFDGNFQAGSSCKAC 356
DB 387 AKKLCRECTAANCKTCD-----QGQCACNDGYKNGDAGSPCHESCKTC 432
QY 357 PANKVOGAVATAGTATILIAQALECPAGIVL---TDGTTSTYKQ-----AASECVKCA 407
DB 433 SA-----GTA-----SDCT-ECPTGKALYKNGDGTGTCGCGCTGTGAGACKTCG 477
QY 408 ANFYTTKQTDWVAGIDTCTSC 428
DB 478 LT-----IDGASYCSEC 489

```

```

RESULT 3
C170_GIALA STANDARD; PRT; 328 AA.
AC P15799;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SURFACE ANTIGEN CRP170 (FRAGMENT).
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 30957 / WB;
RX MEDLINE=88089405; PubMed=3335828;
RA Adam R.D., Aggarwal A., Lal A.A., de la Cruz V.F., McCutchan T.,
RA Nash T.E.;
RT "Antigenic variation of a cysteine-rich protein in Giardia lamblia.";
RL J. Exp. Med. 167:109-118(1988).
CC -!- MISCELLANEOUS: CYSTEINE-RICH, ANTIGENICALLY VARIANT SURFACE
CC PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06741; CAA29916.1; .
DR PIR; S00530;
DR InterPro; IPR002174; Furin-like.
DR SMART; SM00261; FU; 3.
KW Repeat; Antigen.
FT NON-TER 1
FT REPEAT 38 102
FT REPEAT 103 167
FT NON-TER 328
SQ SEQUENCE 328 AA; 33438 MW; 373A697A30EDCA21 CRC64;

Query Match 7.4%; Score 189; DB 1; Length 328;
Best Local Similarity 25.9%; Pred. No. 7.5e-07;
Matches 103; Conservative 36; Mismatches 136; Indels 122; Gaps 28;

QY 36 DDLGTPANVCNCKNF--YNNNAAFVPGASTCTCPQKKGAGQPNPP-----ATANLVT 89
DB 5 DPTGTCSAVDCQGSAGYTTDS--VSDAKECKK-----NAPCTACAGTADKCT 52

QY 90 QC-----NVKCPAGTAIA-----GGA-----TDYAAITTEVCNCRINFYNENAP 128
DB 53 KCDANGAAPYLKKNPSDPTGTCVSADVDCQGSAGYTTDSVSDAKECKK-----NAP 105

QY 129 -NFNAG-ASTCTACPVNRVGGALTAGNAATVAGCNVACPTGT---ALD-----DGVTTD 178
DB 106 CTACAGTADKCTKCAN-----GAAPYLKKNPSDPTGTCVSADVDCQGSAGYTTDD 156

QY 179 YVRSTECVCKRLNFYNGNNGNTPPNPGKSQCTPCPAIRPANVAQTGLNDATITACQN 238
DB 157 SVSDAKECKKA-----EGQKP-NTAGTQCFSC---SDANCERCD-OND--VCARCS 201

QY 239 VACP--DGTISAAGVNNWVAQNTCTNCAPNFYNNAPNPNPGNSTCLPCPANKDYGAEA 296
DB 202 TGAPPENGKCPAA---TPGCHSSCDGCTENAMTNQA-----DKCTGCKEGRYLKPES 250

QY 297 TAG--GAATLAKQCNACPDGTATASATATYVILQTECLNCAANFTYDGNFQAGSSRCK 354
DB 251 AAGSGACLTAEECTSKTHFTREKAGDSKGM-----CLSCS-----DATH---GITGCK 297

QY 355 ACPANKVQGVATAGGTATLAIACALCE-----PAG 385
DB 298 KC-----ALKTLISGEAESTVVCSS-ECTDKRLTPSG 326

```

```

RESULT 4
LMGL_HUMAN STANDARD; PRT; 1609 AA.
AC P11047;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LAMININ GAMMA-1 CHAIN PRECURSOR (LAMININ B2 CHAIN).
GN LAMC1 OR LAMB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91093128; PubMed=1985895;
RA Kallunki T., Ikonen J., Chow L.T., Kallunki P., Tryggvason K.;
RT "Structure of the human laminin B2 chain gene reveals extensive
RT divergence from the laminin B1 chain gene.";
RL J. Biol. Chem. 266:221-228(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88198245; PubMed=3360804;
RA Pikkariainen T., Kallunki T., Tryggvason K.;
RT "Human laminin B2 chain. Comparison of the complete amino acid
RT sequence with the B1 chain reveals variability in sequence homology
RT between different structural domains.";
RL J. Biol. Chem. 263:6751-6758(1988).
RN [3]
RP SEQUENCE OF 1393-1609 FROM N.A.
RX MEDLINE=89169663; PubMed=3234037;
RA Fukushima Y., Pikkariainen T., Kallunki T., Eddy R.L., Byers M.G.,
RA Haley L.L., Henry W.M., Tryggvason K., Shows T.B.;
RT "Isolation of a human laminin B2 (LAMB2) cDNA clone and assignment of
RT the gene to chromosome region 1q25-->q31.";
RL Cytogenet. Cell Genet. 48:137-141(1988).
RN [4]
RP SEQUENCE OF 1282-1609 FROM N.A.
RC TISSUE=Endothelial cells;
RX MEDLINE=92216129; PubMed=1806043;
RA Santos C.L.S., Sabbaga J., Brentani R.;
RT "Differences in human laminin B2 sequences.";
RL DNA Seq. 1:275-277(1991).
CC -!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC -!- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ).
CC LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),
CC LAMININ-6 (K-LAMININ), AND LAMININ-7 (KS-LAMININ).
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -!- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 11 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

	Query Match	7.4%;	Score 188;	DB 1;	Length 1609;	
	Best Local Similarity	23.9%;	Pred. No. 4.3e-06;			
	Matches	99;	Conservative	28;	Mismatches 165;	Indels 122; Gaps 26;
Qy	75	AGQP	PPATANLVTCNVKCPAGTACGATTAAIT	ECVNCRINFYNENAPFNAGA	134	
		: :				
Dd	674	ASAPGGGV	PATWVESCT--CPVG-----YGGQF-----CEWC-LSGYRRPTNLGP-Y	718		
		: :				
Qy	135	STCTACPVNRVGGALTAGNAATI	VAQC--NVACPTGLDDGGVTDTTVYSFTCEVKRLN	192		
Dd	719	SPCVLCACN--GHSETCDPETGCVRDNTAQP	-----HCKCSDG	757		
Qy	193	FYYNGNNGTFFNPGRSQCP	----AIKP-----ANKVAOTLGNDAITTAOCN	238		

FT	CARBOHYD	1908	1908	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	1933	1933	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	2028	2028	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	2111	2111	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	2142	2142	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	2332	2332	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	2533	2533	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	2575	2575	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	2692	2692	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	2812	2812	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	3052	3052	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	3065	3065	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	3144	3144	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	3172	3172	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	3288	3288	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	3292	3292	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	3782	3782	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	4005	4005	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	4136	4136	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	4243	4243	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	4254	4254	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	4335	4335	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	4376	4376	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	4586	4586	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	5136	5136	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	5252	5252	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
SQ	SEQUENCE	5376	5376	AA; 579908 MW; 0E44DB/7DF2A2620 CRC64;		
Query Match 7.4%; Score 188; DB 1; Length 5376;						
Best Local Similarity 21.2%; Pred. No. 1.4e-05;						
Matches 114; Conservative 52; Mismatches 184; Indels 188; Gaps 30;						
QY	17	QIKSANCPCVTETNAGQVDDLGTPANCVN-----CQKNFYNNAAAFVPGASTCTPP-C--	69			
Db	3151	QCONFOCLTKYCK-----DLKDGSSNCTNIPLOCPAHSRYTNC-----LPSCPLCLD	3199			
QY	70	POKKDGAOPNPATANLVTCQNV-----KCPAGTAIAGGATDYAAIIEVCVNCRNF	122			
Db	3200	PEGLCGTSFKVPSTCREGCICOPGYLMHKNK-----VLRIFCGCKNTOGAF	3247			
QY	123	YNENAPNFNAGASTCTACPNRVGGALTAGNATIIYAQC-NVACPTGTALDGVTTDYVR	181			
Db	3248	ISADKTWISRGCTQSCTCP-----AGAI-----HCRNFKCPSGT-----	3281			
QY	182	SFTCEYKCLNRYNGNGNTFPN-----PGKSOCTPC-----PAIKP	219			
Db	3282	-----YCKNGDSSNCTEITLQCPNTSQTDCGLPCVCPSCNRCVETSPSV-P	3329			
QY	220	ANVAATLGNDAI-----TAQCNVACPDGTITSAAGVNNVVAQ-----NTE	260			
Db	3330	SSCREGLCNHGFVSEDKVPRTCQCKDARGAIIIPAG-KTWTSGCTQSCACVGNIIQ	3388			
QY	261	CTN--CAPNFYNNAPNFNGNSTC-----LPCPANKDYGAETA-----GGAATLA	305			
Db	3389	CONFQCPPEY-----CKDNSEGSSTCTKITIQCFAHTQYTSCLPCLDPSELCKDIS	3445			
QY	306	KOCNIACPDGTAIASGATNYVILQTECLNCAANFYFDGNNFQAGSSRCKACAPANKVQGV	365			
Db	3446	PKVSTCKEGVCOSG-----YVLNSDKCVLRA-----BCDCKDAQAL	3484			
QY	366	ATAGGTATL-----IAQC-ALECPAGTIVLDGTTSTYKQASCEVK-----CAA	408			
Db	3485	IPAGKTWTSPTGCTQSCACMGAVQCQSQCPPGTYCKDN-----EDGNSNGAKITLOC	3539			
QY	409	NYFTTKQTDWAGIDTC-TSC--NKKLTSGAEANLPSAKNIOCDFANFLISLLLI	463			
Db	3540	HSLFTN-----CLPPCLPCLDPDGLCKGASPKVPSTCKEGCICQSGYVLSNNKCLL	3591			
RESULT	6					
PCK5_BRACL						
ID		STANDARD;	PRT; 1696	AA.		

AC	Q9NJ15; Q9NJ16; Q9NJ14;	
DT	20-AUG-2001 (Rel. 40, Created)	
DT	20-AUG-2001 (Rel. 40, Last sequence update)	
DT	20-AUG-2001 (Rel. 40, Last annotation update)	
DE	PROTEIN CONVERTASE SUBTILISIN/KEXIN TYPE 5 PRECURSOR (EC 3.4.21.-)	
DE	(PROTEIN CONVERTASE PC6-LIKE) (APC6).	
GN	PC6.	
OS	Branchiostoma californiensis (California lancelet) (Amphioxus).	
OC	Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;	
OC	Branchiostoma.	
OX	NCBI_TaxID=7738;	
RN	[1]	
RP	SEQUENCE FROM N.A. (ISOFORM A; ISOFORM B AND ISOFORM C).	
RX	MEDLINE=20175281; PubMed=10708868;	
RA	Oliva A.A. Jr., Chan S.J., Steiner D.F.;	
RT	"Evolution of the prohormone convertases; Identification of a	
RT	homologue of PC6 in the protochordate amphioxus.";	
RL	Biochim. Biophys. Acta 1477:338-348(2000).	
CC	-!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY	
CC	WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE	
CC	OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (3Y SIMILARITY).	
CC	-!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR	
CC	PROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG- -ZAA BONDS, WHERE XAA	
CC	CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS	
CC	-!- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.	
CC	ISOFORM B IS A TYPE I MEMBRANE PROTEIN.	
CC	-!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A, B (SHOWN HERE) AND C; ARE	
CC	PRODUCED BY ALTERNATIVE SPLICING.	
CC	-!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE	
CC	ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC	
CC	RETICULUM.	
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE	
CC	SUBTILASE FAMILY.	
CC	-!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; AF184615; AAF26300.1; -	
DR	EMBL; AF184616; AAF26301.1; -	
DR	EMBL; AF184617; AAF26302.1; -	
DR	InterPro; IPR000561; EGF-like.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR002884; P_domain.	
DR	InterPro; IPR000209; Peptidase_S8.	
DR	ProDom; PD000717; P_domain; 1.	
DR	SMART; SM00181; EGF; 2.	
DR	SMART; SM00261; FU; 17.	
DR	PROSITE; PS00136; SUBTILASE ASP; FALSE_NEG.	
DR	PROSITE; PS00137; SUBTILASE HIS; 1.	
DR	PROSITE; PS00138; SUBTILASE SER; 1.	
KW	Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;	
KW	Cleavage on pair of basic residues; Repeat; Alternative splicing;	
KW	Transmembrane.	
FT	SIGNAL	1 25 POTENTIAL.
FT	PROPEP	26 110 POTENTIAL.
FT	CHAIN	111 1696 PROTEIN CONVERTASE SUBTILISIN/KEXIN
FT		TYPE 5.
FT	DOMAIN	111 1618 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1619 1639 POTENTIAL.
FT	DOMAIN	1640 1696 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	111 488 CATALYTIC.
FT	DOMAIN	496 637 HOMO B.
FT	DOMAIN	664 1649 CYS-RICH MOTIF (CRM) REGION.
FT	SITE	110 111 CLEAVAGE (AUTO-) (BY SIMILARITY).
FT	ACT_SITE	192 192 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	233 233 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	407 407 CHARGE RELAY SYSTEM (BY SIMILARITY).

```
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 885 885 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1259 1323 DDILRGECITSCGPGEMDRREKKKACHPTKCEKSDVEY
FT DDTCTACNDGFLTLTDASSCEAGCP -> AENQNOASFQCPFA
FT PREVSVALAELALHRLYSLTDVPPQSNPPDVLGADRRL
FT TTATSAAGRA (IN ISOFORM C).
FT VARSPLIC 1324 1696 CHPTCKECSDVDYDCTACNDGFLTLTDASSCEAGCPGQFL
FT VARSPLIC 1288 1343 HUGDCKSHRECKTC -> IARCVDRRDRSCWDLVLRNFC
FT VRYFVKRCGCKKLYMEDRPMRRGSSQPTOGRN (IN
FT ISOFORM A).
FT VARSPLIC 1344 1696 MISSING (IN ISOFORM A).
FT SEQUENCE 1696 AA; 188410 MW; 281CBEL784257CBD CRC64;

Query Match 7.3%; Score 185.5; DB 1; Length 1696;
Best Local Similarity 20.4%; Pred. No. 6.9e-06;
Matches 101; Conservative 43; Mismatches 189; Indels 161; Gaps 25;

QY 18 IKSANCPVGTETAG-----QVDDLGTPANCVCOKNFYNNAAAFVPGASTCT 67
DB 1048 VRTNCPSFTYDDODRECPCHDNEACDGPNNQNCSCKEGYKT-----PDGCS--T 1100
QY 68 PCPQ---KIDAGAPNP-----PATANLY-----TQCNVCKPAGTAIAGG 104
DB 1101 GCPNRYKDDTNKEKPCDSCFTCSGPASFCLSCADGDFLHSSCRSTCPAG--FIGN 1158
QY 105 ATDYAALITECVNCRINFYNENAPNFNAGASTCTACPNRVGGALTAGNAATVACQNV 164
DB 1159 AESHECVESSC-----EQQYYSSETGCEDCPYN-----CRACDNDGDCAEC 1201
QY 165 CPTGTALD-----DGVTTDYVRSFTECVKCRNFYNNNGNTNPFNPKSQCTPC 214
DB 1202 APTIYVVDGRCPEETCEDEGEYDRDRAE--LSCH-----FCHOSCKTC 1245
QY 215 PALKPANVAQAATLGNDATTIACQNVACPGDGTISAAGVNNVAQNTNCTC--TNCAPNFY--NNN 272
DB 1246 SG--PSD-----TDCDCKGDDTILDRG-----ECITSCGPGVEMDRR 1281
QY 273 APNFNPGNSTCLPCPANKDYGAETA---GGAATLAKOCNIACPDGTATASCATNYVILQ 329
DB 1282 EKKCKACHPTCKEC--SDEYDDTCTACNDGFLTLTDASSCEAGCPGQFLHHG----- 1331
QY 330 TECLNCAANFYDGNFNQAGSSRCKACPAKQVGAATAGGTATLIAOCALCECPAGTVLT 389
DB 1332 -BCDSC-----HRECKTCGPHDNCILSCOPGSLYNDQCSHCPCGTF-- 1374
QY 390 DGTSTYKQAASE---CVKCAANFYTKQTDVAGIDTCTSC-----NKKLTSGA 436
DB 1375 ---EETIEDDSGETVLQCRCLHVNCKTCHG-----EGEEDCMECANDIKYKODGRCVTECQ 1427
QY 437 EANLPESAKKNQOC 450
DB 1428 EGHYPDLTNECQOC 1441

RESULT 7
VS41_GIALA
ID VS41_GIALA STANDARD; PRT; 687 AA.
AC P92127;
DT 20-AUG-2001 (Rel. 40, Created)
DE 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O2-4A1;
RX MEDLINE=97321554; PubMed=9178264;
RA Papanastasiou P., Bruderer T., Li Y., Bommeili C., Koehler P.;
```

```
RT RT Primary structure and biochemical properties of a variant-specific
RL RL surface protein of Giardia.;
RN RN Mol. Biochem. Parasitol. 86:13-27(1997).
RX CHARACTERIZATION.
RA MEDLINE=97233006; PubMed=9078242;
RA Papanastasiou P., McConville M.J., Ralton J., Koehler P.;
RT "The variant-specific surface protein of Giardia, VSP4A1, is a
RL glycosylated and palmitoylated protein.";
RL Biochem. J. 322:49-56(1997).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
CC PLASMA MEMBRANE.
CC -!- PTM: O-GLYCOSYLATED. THE MAJOR GLYCAN IS A TRISACCHARIDE WITH GLC
CC AT THE REDUCING TERMINUS.
CC -!- PTM: PALMITOYLATED.
CC -!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; Z83743; CAB06038.1; -
DB GlycoSuiteDB; P92127; -
DB InterPro; IPR000561; EGF-like.
DB InterPro; IPR002174; Furin-like.
DB SMART; SM00181; EGF; 2.
DB SMART; SM00001; EGF_Like; 1.
DB SMART; SM00261; FU; 3.
DB Antigen; Repeat; Transmembrane; Glycoprotein; Lipoprotein; Palmitate;
KW SIGNAL. 1 14 POTENTIAL.
FT CHAIN 15 687 VARIANT-SPECIFIC SURFACE PROTEIN VSP4A1.
FT DOMAIN 15 660 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 661 681 POTENTIAL.
FT DOMAIN 682 687 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 687 AA; 70857 MW; D892F675D626D7EC CRC64;

Query Match 7.3%; Score 184.5; DB 1; Length 687;
Best Local Similarity 21.1%; Pred. No. 3.3e-06;
Matches 119; Conservative 54; Mismatches 214; Indels 177; Gaps 29;

QY 12 SLFINQIKSANCPVGTET-----NTAGQVDDLGTPANCVCOKNFYNNAAAFVPGASTCT 67
DB 124 SOYIFQNKATPSEKSGSECLCWDITDRNGYGV-ANCATC-----TAPASSTGPATCT 175
QY 68 PCP-----QKDKAGAPNPATANLVTCNVKCPAGTAIAGGATDYAALITECVNCRINFY 123
DB 176 ECMAGTYKKSD-----TEC-AACHSDCATCSGEAN-----NQCTSCETGKY 215
QY 124 -----NENAPFN-----AGASTCTACPNRVGGALTAGNAATVACQNVACPTG----- 168
DB 216 LKSNQCVKEKTCNTNHYPPDDTSMTCVACTVLDANCATCSFSDATAGKGC-LTCNSKNIPR 274
QY 169 TALDDGVTTDYVRSFTECVKCRNFYNNNGNTNPFNPKSQCTPCPAIKPANVAQAATLG 228
DB 275 TTL-DGTSTCVENSYAGC-----QGADNELFMKEDQSACLLCGDTKEASNDKG-VA 323
QY 229 NDATITTAQCN-----VACPDGTISAAG-VNNVVAQN-----TECTNCAPNFYNN 271
DB 324 NCRTCTKNANDSPPTCTACLDGYFLERGSGTCTTCADNCATCSSEATTEDCKICKAGFF-- 381
QY 272 NAPNFNPGNSTCLPCPANKDYG---AEATAGGAATL-AKQC-----NIACPDG 315
DB 382 ---LASPEGCKCISCSDTNNNGIDGCAECTKEPAGPLKCTKCPNRRKPAGTSDNYTCTEK 438
QY 316 TAI-----ASGATNYVILQTE-----C 332
DB 439 TCEDPTVCGGTSGACDAIVIDANGKEHYVCSYCGETNKKFPIDGLCTDNKGTNAGCTDHTC 498
```


333 LNCAANFYEDGNFNQAGSSKRCACPA NKVGAVATAGGTATLIA-----OC 378
 ||| : : : ||| |
 499 SYCAAGFEL---YMGCCYKIDITPGSYMCKSRTTATGVDTPNANRRFFVVPKAISAQS 554
 - - - Y M G C C Y K I D I T P G S Y M C K S R T T A T G V D T P N A N R R F F V V P K A I S A Q S
 379 ALBC--PAGTVLTDGTSIYYKQAASCEVCRAANFYTKQTDMWAGDTCTSCNKKITSGA 436
 ||| : : : ||| |
 555 VLACGNPLGTI-AGGNAYGVGECSOCTAPDA-----RADGGMAVATCTACEDGKKPG- 606
 - - - - - Y A G N A Y G V G E C S O C T A P D A -----R A D G G M A V A T C T A C E D G K K P G -
 437 EANLPESAKNKNIQCDEANFLSLISL 460
 : | : | | | : :
 607 -----KSgtgcvacpdancskctm 625

```

RESULT      8
UR2_DROME  UR2_DROME    STANDARD;          PRT; 1680 AA.

C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C
D       D       D       D       D       D       D       D       D       D       D       D       D       D       D       D       D       D
E       E       E       E       E       E       E       E       E       E       E       E       E       E       E       E       E       E
F       F       F       F       F       F       F       F       F       F       F       F       F       F       F       F       F       F
G       G       G       G       G       G       G       G       G       G       G       G       G       G       G       G       G       G
H       H       H       H       H       H       H       H       H       H       H       H       H       H       H       H       H       H
I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I
J       J       J       J       J       J       J       J       J       J       J       J       J       J       J       J       J       J
K       K       K       K       K       K       K       K       K       K       K       K       K       K       K       K       K       K
L       L       L       L       L       L       L       L       L       L       L       L       L       L       L       L       L       L
M       M       M       M       M       M       M       M       M       M       M       M       M       M       M       M       M       M
N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N
O       O       O       O       O       O       O       O       O       O       O       O       O       O       O       O       O       O
P       P       P       P       P       P       P       P       P       P       P       P       P       P       P       P       P       P
Q       Q       Q       Q       Q       Q       Q       Q       Q       Q       Q       Q       Q       Q       Q       Q       Q       Q
R       R       R       R       R       R       R       R       R       R       R       R       R       R       R       R       R       R
S       S       S       S       S       S       S       S       S       S       S       S       S       S       S       S       S       S
T       T       T       T       T       T       T       T       T       T       T       T       T       T       T       T       T       T
U       U       U       U       U       U       U       U       U       U       U       U       U       U       U       U       U       U
V       V       V       V       V       V       V       V       V       V       V       V       V       V       V       V       V       V
W       W       W       W       W       W       W       W       W       W       W       W       W       W       W       W       W       W
X       X       X       X       X       X       X       X       X       X       X       X       X       X       X       X       X       X
Y       Y       Y       Y       Y       Y       Y       Y       Y       Y       Y       Y       Y       Y       Y       Y       Y       Y
Z       Z       Z       Z       Z       Z       Z       Z       Z       Z       Z       Z       Z       Z       Z       Z       Z       Z
```

01-APR-1993 (Rel. 25, Created)
 01-APR-1993 (Rel. 25, Last sequence update)
 15-JUL-1998 (Rel. 36, Last annotation update)
 FURIN-LIKE PROTEASE 2 PRECURSOR (EC 3.4.21.75) (FURIN 2).
 FUR2.
 Drosophila melanogaster (Fruit fly).
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Epheuroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 [!]
 SEQUENCE FROM N.A.
 MEDLINE=92381036; PubMed=1512259;
 Roelbreck A.J.M., Creemers J.W.M., Pauli I.G.L., Kurzik-Dumke U.,
 Rentrop M., Gateff E.A.F., Leunissen J.A.M., van de Ven W.J.:
 "Cloning and functional expression of Furin2, a subtilisin-like
 proprotein processing enzyme of *Drosophila melanogaster* with multiple
 repeats of a cysteine motif.";
 J. Biol. Chem. 267:17208-17215(1992).
 FUNCTION: FURIN IS LIKELY TO REPRESENT THE UBQUITOUS ENDOPEPTIDASE
 ACTIVITY WITHIN CONSTITUTIVE SECRETORY PATHWAYS AND CAPABLE OF
 CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (BY SIMILARITY).
 !- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
 PROPROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA
 CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS. RELEASES ALBUMIN,
 COMPLEMENT COMPONENT C3 AND VON WILLEBRAND FACTOR FROM THEIR
 RESPECTIVE PRECURSORS.
 !- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 SUBTILASE FAMILY. FURIN SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch)

EMBL; M94375; AAA28551.1; -.
 FIR; A4343; A43434.
 HSP; Q99405; LMPT.
 FlyBase; FBgn0004598; Fur2.
 InterPro; IPRO00561; EGF-like.
 InterPro; IPRO02174; Furin-like.
 InterPro; IPRO02884; P_domain.
 InterPro; IPRO00209; Peptidase_S8.
 Pfam; PF00757; Furin-like; 1.
 Pfam; PF01483; P; 1.
 Pfam; PF00082; Peptidase_S8; 1.
 PRINTS; PR00723; SUBTILISIN.
 ProDom; PD000717; P_domain; 1.
 SMART; SM00181; EGF; 1.
 SMART; SM00261; FU; 10.

DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR

QY 398 -QAASECVKCAANFYTTKQTDWVAGIDTCTSCN 429
 Db 1428 DQLNSQCVSCCQNTLAETSSAA----CNCD 1456

RESULT 9
 PKC5_MOUSE STANDARD; PRT: 1877 AA.
 AC 004592; Q62040;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROTEIN CONVERTASE SUBTILISIN/KEXIN TYPE 5 PRECURSOR (EC 3.4.21.-)
 DE (PROTEIN CONVERTASE PC5) (SUBTILISIN/KEXIN-LIKE PROTEASE PC5)
 DE (CONVERTASE PC5) (PC6) (SUBTILISIN-LIKE PROTEIN CONVERTASE 6)
 DE (SPC6).
 GN PCSK5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PC5B).
 RC STRAIN=ICR; TISSUE=Intestine;
 RX MEDLINE=93327934; PubMed=8335106;
 RA Nakagawa T., Murakami K., Nakayama K.;
 RT "Identification of an isoform with an extremely large Cys-rich region
 of PC6, a kex2-like processing endoprotease.";
 RL FEBS Lett. 327:165-171(1993).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM PC5A).
 RC TISSUE=Brain, and Intestine;
 RX MEDLINE=933224489; PubMed=8468318;
 RA Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K.,
 RA Nakayama K.;
 RT "Identification and functional expression of a new member of the
 mammalian kex2-like processing endoprotease family: its striking
 structural similarity to PACE4.";
 RL J. Biochem. 113:132-135(1993).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM PC5A).
 RC TISSUE=Adrenal cortex;
 RX MEDLINE=93342056; PubMed=8341687;
 RA Lusson J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;
 RT "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a
 candidate proprotein convertase expressed in endocrine and
 nonendocrine cells.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).
 RN [4]
 RP PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.
 RX MEDLINE=97103178; PubMed=8947550;
 RA De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K.,
 RA Bendayan M., Seidah N.G.;
 RT "The isoforms of proprotein convertase PC5 are sorted to different
 subcellular compartments.";
 RL J. Cell Biol. 135:1261-1275(1996).
 RN [5]
 RP DEVELOPMENTAL EXPRESSION.
 RX MEDLINE=96293359; PubMed=8698813;
 RA Constam D.B., Calton M., Robertson E.J.;
 RT "SPC6, a PC6, and the novel protease SPC7 are coexpressed with bone
 morphogenetic proteins at distinct sites during embryogenesis.";
 RL J. Cell Biol. 134:181-191(1996).
 RN [6]
 RP DEVELOPMENTAL EXPRESSION.
 RX MEDLINE=97436919; PubMed=9291583;
 RA Rancourt S.L., Rancourt D.E.;
 RT "Murine subtilisin-like proteinase SPC6 is expressed during embryonic
 implantation, somitogenesis, and skeletal formation.";
 RL Dev. Genet. 21:75-81(1997).
 CC -1- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
 WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
 OF CLEAVAGE AT THE R(X/K/R)R CONSENSUS MOTIF. MAY BE RESPONSIBLE

CC FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED
 CC IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION
 CC OF GROWTH FACTORS.
 CC -1- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
 CC PROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA
 CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
 CC -1- SUBCELLULAR LOCATION: PC5A IS SECRETED THROUGH THE REGULATED
 CC SECRETORY PATHWAY. PC5B IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO
 CC A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH
 CC EARLY ENDOSOMES.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS: PC5B/LONG (SHOWN HERE)
 CC AND PC5A/SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: PC5A IS EXPRESSED IN MOST TISSUES BUT IS MOST
 CC ABUNDANT IN THE INTESTINE AND ADRENALS. PC5B IS EXPRESSED IN THE
 CC INTESTINE, ADRENALS AND LUNG BUT NOT IN THE BRAIN.
 CC -1- DEVELOPMENTAL STAGE: WEAKLY EXPRESSED THROUGHOUT THE EMBRYO.
 CC EXCEPT IN THE DEVELOPING NERVOUS SYSTEM, THE RIBS AND THE LIVER,
 CC BUT MARKEDLY UPREGULATED AT DISCRETE SITES DURING DEVELOPMENT. AT
 CC E6.5, PROMINENT EXPRESSION OBSERVED IN DIFFERENTIATED DECIDUA. AT
 CC E7.5, INTENSE EXPRESSION IN EXTRAEMBRYONIC ENDODERM, AMNION AND
 CC NASCENT MESODERM. AT 8.5, ABUNDANT EXPRESSION IN SOMITES AND YOLK
 CC SAC FOLLOWED BY A CONFINATION TO DERMATOTOME COMPARTMENT BETWEEN
 CC E9.5 AND E11.5. ABUNDANT EXPRESSION IN AER (THICKENED ECTODERMAL
 CC CELLS OF LIMB BUDS). AT E12.5, EXPRESSION IN THE LIMBS IS CONFINED
 CC TO THE CONDENSING MESENCHYM SURROUNDING THE CARTILAGE. AT THIS
 CC STAGE, STRONG EXPRESSION ALSO DETECTED IN VERTEBRAL AND FACIAL
 CC CARTILAGE PRIMORDIA AND IN THE MUSCLE OF THE TONGUE. AT E16.5,
 CC ABUNDANT EXPRESSION IN EPITHELIAL CELLS OF THE INTESTINAL VILLI.
 CC ISOFORM A IS MOST ABUNDANT AT ALL STAGES BUT SIGNIFICANT LEVELS OF
 CC ISOFORM B OCCUR AT E12.5.
 CC -1- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
 CC ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC
 CC RETICULUM.
 CC -1- DOMAIN: AC 1 AND AC 2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN
 CC SORTING INFORMATION. AC 1 DIRECTS TGN LOCALIZATION AND INTERACTS
 CC WITH THE TGN SORTING PROTEIN PACS-1.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D17583; BAA04507.1; -;
 CC EMBL: D12619; BAA02143.1; -;
 CC EMBL: L14932; AAA74636.1; -;
 CC PIR: JX0248; JX0248.
 CC PIR: A48225; A48225.
 CC HSP: Q99405; 1MPT.
 CC MEROPS: S08.076; -;
 CC MGD: MGI:97515; Pcsk5
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR002174; Furin-like.
 CC InterPro: IPR002884; P_domain.
 CC InterPro: IPR000209; Peptidase_s8.
 CC Pfam: PF01483; P; 1.
 CC Pfam: PF00082; Peptidase_S8; 1.
 CC PRINTS: PR00723; SUBTILISIN.
 CC ProDom: PD000717; P_domain; 1.
 CC SMART: SM00181; EGF; 3.
 CC SMART: SM00001; EGF_like; 2.
 CC SMART: SM00261; FU; 22.
 CC PROSITE: PS00136; SUBTILASE ASP; 1.
 CC PROSITE: PS00137; SUBTILASE HIS; 1.
 CC PROSITE: PS00138; SUBTILASE SER; 1.
 CC HydroLase: Serine protease; Glycoprotein; Zymogen; Signal;
 CC Cleavage on pair of basic residues; Repeat; Alternative splicing;
 CC Transmembrane.

```

Query Match          7.1%; Score 179.5; DB 1; Length 1877;
Best Local Similarity 20.3%; Pred. No. 2.le-05;
Matches 96; Conservative 47; Mismatches 166; Indels 163; Gaps 25;

QY 42 ANCVNCQKNFYNNAAAFVPGASTCTPCQKQKADGAQNPATANLVITQCNVKCPAGT-- 99
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1214 SSCKTC-----NGSLCASCPGTMYLWQ-----ACVSPCQGTWP 1248
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 100 ATAGGATDYAIIITCVNCRINFYNENAPNFWAGASTCTAC---PVNRYGGALTAGNAAT 156
   : : : : : : | : | : | : | : | : | : | : | : | : | : | : |
Db 1249 SVTSGCEKCS--EDCVSC-----SGADLCOQCLSPDNTL--LLHEG---- 1287
   : : : : : : | : | : | : | : | : | : | : | : | : | : | : |

QY 157 IVAOCNVACPTGALTDDGGVYTDYVRSFTCVKCRLNFYNGNGNTPNPFGSKQCTPCPA 216
   : : | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 1288 ---RCVHSCPEGYAKGV-----CEHC-----SSPCKTCBGNATSCNS 1323
   : : | : | | | | | | | | | | | | | | | | | | | | | |

QY 217 IKPANVAQATLGNATITIAQCNVACPDGTISAGVNNVQAQNT-----CTNCAPNY 269
   : : | : | | | : | : | : | : | : | : | : | : | : | : | : |
Db 1324 CE-----GDFVLHDGVCKWKTCEKHAVEGVCKHCPCRCQCIHBEKTCRCMPDF 1374
   : : | : | | | | | | | | | | | | | | | | | | | | | |

QY 270 NNN-----APNFPNGNSTCLPCPAN-----KDYGAETAGGAATLEAK--QCNIACPD 314
   : : | : | | | | | | | | | | | | | | | | | | | | | |
Db 1375 LYNDMCHRSCPFSFYDPMRCQVPCCHKNCLECGPKEDDKCVKADTSKALHNGLCLECFE 1434
   : : | : | | | | | | | | | | | | | | | | | | | | | |

QY 315 GT-----AATASGA-----TNYILOTECL--NCAANFYFDGNFFOA 348
   : : | : | | | | | | | | | | | | | | | | | | | | | |
Db 1435 GTYKEBDECDPCPESCLICSSAWTCLACREGFTVVHVDCVTAPECAAVEYWD-----E 1489
   : : | : | | | | | | | | | | | | | | | | | | | | | |

QY 349 GSSRCRKAC--PANKVOG-----AVATAGGTATLIIAOCALCEPAG-----TVLTDTGT 392
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1490 GSHRCQCPCHKKSRCSGPSDQCYTCPRETFLLNTTTCVKCEPEGYHTDKDQOCVLCSS 1549
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 393 TSYIKQAAS--ECVKCAANFYTTKQ-----TDMVAGIDT---CTSCNKKLTS 434
   : : | : | | : : | : | : | : | : | : | : | : | : | : | : |
Db 1550 CRTCEGPHSMQCLSRPGWFLQKCECLLCROCRGYGYESTSGRECKDKSCKS 1601
   : : | : | | : : | : | : | : | : | : | : | : | : | : | : |

```

Query Match	7.0%;	Score 179;	DB 1;	Length 2715;
Best Local Similarity	23.7%;	Pred. No. 3.2e-05;		
Matches 125;	Conservative 38;	Mismatches 198;	Indels 156;	Gaps 32;
QY	20	SANCPVGTETNTAGQVDDLGTPANCVNCOKFEYNNAA--AFVPGA-----STC-----TP 68		
Db	970	AANCFFRSSASGTAGYC-----AMNTNCS---VTSAAECAVFGLTGLDHSKCOLYHSS 1020		
QY	69	CPQKQD-AGAAQ-----PNPPATANLVTCQNVKC-----PAGTAIAG---G 104		
Db	1021	CTSILKDGTCQEYKTTCSGYAATNNCATSGQKCFDVECLRFSNCASITGTGLTAICG 1080		
QY	105	ATDYAAII-----TEC-----VNCRINFINENAPNFNAGASTCTACPVNRVGGALTAGNAAAT 156		
Db	1081	TYDAGCVYANVNGTACQEKLATCDL-YLTQNSCTSAATAADK-----AWSGTACL 1131		

QY 157 IVAOCNVACP--TGATLDGCVTTDYVRSFTECVKCRLENYFNNNGNTNPNFGKSOQ----- 211
 Db 1132 AVTTGTHCPYVGTGTDLICAA-----NANCTANKAGTACQEKKATCNLY 1179
 QY 212 ---TPCPAIPKPNVAQATLGNDAITIAQCNVACPDCGTISAAGVNNWV--AQNTCTNCAP 266
 Db 1180 TTEATCTSAATAADKCAWSGAACLAIVTVATECAAYVGTGLTDLICAYNANCT----- 1235
 QY 267 NFYNNAFPNPNSTCLPCPANKD--YGAET--AGGAATLAKQC---NIACPDTGTAIA- 319
 Db 1236 -----ANKAGTACQEKKATCNLYTTEATCTSAATAADKCAWSGAACLAIVTVAT 1286
 QY 320 -----SGATNVVILQTECLNCAANFYDGNFQAGSSRCRCACPAK----- 360
 Db 1287 ECAYVTGTGLTN-AICAAVYNNCTAN-----RAGT-----ACQEKKATCNLYTTEATC 1333
 QY 361 -VOGAVATA-----GGTATLQAQCALECP--AGTVLTDGTTSTY-----KQAASEC-- 403
 Db 1334 STSAAATAADKCAWSGAACLAIVTVATECAAYVGTGLTNAICAAVYNNCTANKAGTACQE 1393
 QY 404 VKCAAFNYTTKOTDWDVAGIDTCTSCNKKLTSCAEANLPESAKKNQOC 450
 Db 1394 KKATCNLYTTEAT-----CS---TSAAATAADKCAWSGAAC 1426

RESULT 11
 LMGL_MOUSE
 ID LMGL_MOUSE STANDARD; PRT: 1607 AA.
 AC P02468;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE LAMININ GAMMA-1 CHAIN PRECURSOR (LAMININ B2 CHAIN).
 GN LAMC1 OR LAMC-1 OR LAMB-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88059118; PubMed=3680290;
 RA Sasaki M, Yamada Y.;
 RT "The laminin B2 chain has a multidomain structure homologous to the
 RT B1 chain.";
 RL J. Biol. Chem. 262:17111-17117(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89000737; PubMed=3167041;
 RA Durkin M.E., Bartos B.B., Liu S.-H., Phillips S.L., Chung A.E.;
 RT "Primary structure of the mouse laminin B2 chain and comparison with
 RT laminin B1.";
 RL Biochemistry 27:5198-5204(1988).
 RN [3]
 RP SEQUENCE OF 1-239 FROM N.A.
 RX MEDLINE=88228071; PubMed=2836421;
 RA Ogawa K., Burbello P.D., Sasaki M., Yamada Y.;
 RT "The laminin B2 chain promoter contains unique repeat sequences and
 RT is active in transient transfection.";
 RL J. Biol. Chem. 263:8384-8389(1988).
 RN [4]
 RP SEQUENCE OF 1391-1607 FROM N.A.
 RX MEDLINE=85051302; PubMed=6209134;
 RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;
 RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of
 RT coiled-coil alpha-helix.";
 RL EMBO J. 3:2355-2362(1984).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 771-932.
 RX MEDLINE=96196434; PubMed=8648630;
 RA Stetefeld J., Mayer U., Timpl R., Huber R.;
 RT "Crystal structure of three consecutive laminin-type epidermal growth
 RT factor-like (LE) modules of laminin gamma1 chain harboring the

RT nidogen binding site.";
 RL J. Mol. Biol. 257:644-657(1996).
 RN [6]
 RP STRUCTURE BY NMR OF 824-881.
 RX MEDLINE=96196435; PubMed=8648631;
 RA Baumgartner R., Csisch M., Mayer U., Poeschl E., Huber R.,
 RA Timpl R., Holak T.A.;
 RT "Structure of the nidogen binding LE module of the laminin gamma1
 RT chain in solution.";
 RL J. Mol. Biol. 257:658-668(1996).
 CC -!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
 CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -!- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ),
 CC LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),
 CC LAMININ-6 (K-LAMININ), AND LAMININ-7 (KS-LAMININ).
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR
 CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT).
 CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -!- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
 CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -!- SIMILARITY: CONTAINS 11 LAMININ EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X05211; CAA28838.1; -;
 DR EMBL; J03484; AAA39405.1; -;
 DR EMBL; J02930; AAA39408.1; -;
 DR EMBL; J03749; AAA39409.1; -;
 DR PIR; A28469; MMSB2.
 DR PDB; IKLO; 20-AUG-97.
 DR PDB; ITLE; 12-FEB-97.
 DR MGD; MGI:99914; Lamc1.
 DR InterPro; IPR000122; Chemotaxis_transducer.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001886; LamNT.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00052; laminin_B; 1.
 DR Pfam; PF00053; laminin_EGF; 10.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR PRODOM; PD002082; LamNT; 1.
 DR PRODOM; PD003031; Laminin_B; 1.
 DR SMART; SM00180; EGF_Lam; 9.
 DR SMART; SM00001; EGF_like; 1.
 DR SMART; SM00281; LamB; 1.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00022; EGF_1; 8.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 10.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
 FT SIGNAL 1 33
 FT CHAIN 34 1607 LAMININ GAMMA-1 CHAIN.
 FT DOMAIN 34 283 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 284 502 4.5 X LAMININ EGF-LIKE REPEATS (DOMAIN
 FT V).
 FT DOMAIN 284 339 LAMININ EGF-LIKE 1.

Biochem. J. 276:369-379(1991).
 [3]
 RN SEQUENCE OF 2397-3072 FROM N.A.
 RP MEDLINE=89280632; PubMed=2733383;
 RX Olsen D., Nagayoshi T., Fazio M., Peltonen J., Jaakkola S.,
 RA Sanborn D., Sasaki T., Kuivaniemi H., Chu M.L., Deutzmann R.,
 RA Timpl R., Utito J.;
 RT "Human laminin: cloning and sequence analysis of cDNAs encoding A, B1
 RT and B2 chains, and expression of the corresponding genes in human
 RT skin and cultured cells";
 RL Lab. Invest. 60:772-782(1989).
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
 CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC THE ALPHA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ) AND
 CC LAMININ-3 (S'-LAMININ).
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT).
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -1- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X58531; CAA41418.1; .
 CC PIR; S14458; S14458.
 CC HSP; P02468; 1TLE.
 CC MIM; 150320; .
 CC InterPro; IPR000561; EGF-like.
 CC InterPro; IPR001886; LamNT.
 CC InterPro; IPR000034; Laminin.B.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR001791; Laminin_G.
 CC Pfam; PF00052; laminin_B; 2.
 CC Pfam; PF00053; laminin_EGF; 15.
 CC Pfam; PF00054; laminin_G; 5.
 CC Pfam; PF00055; laminin_Nterm; 1.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC PRODOM; PD02082; LamNT; 1.
 CC PRODOM; PD03031; Laminin_B; 2.
 CC SMART; SM00180; EGF_Lam; 14.
 CC SMART; SM00001; EGF_like; 1.
 CC SMART; SM00281; LamB; 2.
 CC SMART; SM00282; LamG; 5.
 CC SMART; SM00136; LamNT; 1.
 CC PROSITE; PS00022; EGF_1; 11.
 CC PROSITE; PS01186; EGF_2; 2.
 CC PROSITE; PS01248; LAMININ_TYPE_EGF; 15.
 CC Glycoproteins; Basement membrane; Extracellular matrix; Coiled coil;
 CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 CC SIGNAL 1 17 POTENTIAL.
 CC CHAIN 18 3075 LAMININ ALPHA-1 CHAIN.
 CC DOMAIN 18 269 LAMININ N-TERMINAL (DOMAIN VI).
 CC DOMAIN 270 516 4.5 X LAMININ EGF-LIKE REPEATS (DOMAIN
 CC V).
 CC DOMAIN 270 326 LAMININ EGF-LIKE 1.
 CC DOMAIN 327 396 LAMININ EGF-LIKE 2.

FT DOMAIN 397 453 LAMININ EGF-LIKE 3.
 FT DOMAIN 454 502 LAMININ EGF-LIKE 4.
 FT DOMAIN 503 512 LAMININ EGF-LIKE 5 (N-TERMINAL).
 FT DOMAIN 517 708 LAMININ DOMAIN IV 1 (DOMAIN IV B).
 FT DOMAIN 709 1159 9 X LAMININ EGF-LIKE REPEATS (DOMAIN
 III B).
 FT DOMAIN 709 741 LAMININ EGF-LIKE 5 (C-TERMINAL).
 FT DOMAIN 742 790 LAMININ EGF-LIKE 6.
 FT DOMAIN 791 848 LAMININ EGF-LIKE 7.
 FT DOMAIN 849 901 LAMININ EGF-LIKE 8.
 FT DOMAIN 902 950 LAMININ EGF-LIKE 9.
 FT DOMAIN 951 997 LAMININ EGF-LIKE 10.
 FT DOMAIN 998 1043 LAMININ EGF-LIKE 11.
 FT DOMAIN 1044 1089 LAMININ EGF-LIKE 12.
 FT DOMAIN 1090 1149 LAMININ EGF-LIKE 13.
 FT DOMAIN 1150 1159 LAMININ EGF-LIKE 14 (N-TERMINAL).
 FT DOMAIN 1160 1361 LAMININ DOMAIN IV 2 (DOMAIN IV A).
 FT DOMAIN 1362 1555 3.5 X LAMININ EGF-LIKE REPEATS (DOMAIN
 III A).
 FT DOMAIN 1362 1402 LAMININ EGF-LIKE 14 (C-TERMINAL).
 FT DOMAIN 1403 1451 LAMININ EGF-LIKE 15.
 FT DOMAIN 1452 1508 LAMININ EGF-LIKE 16.
 FT DOMAIN 1509 1555 LAMININ EGF-LIKE 17.
 FT DOMAIN 1556 2125 DOMAIN II AND I.
 FT DOMAIN 2126 2075 5 X LAMININ G-LIKE REPEATS (DOMAIN G).
 FT DOMAIN 2140 2327 LAMININ G-LIKE 1.
 FT DOMAIN 2328 2509 LAMININ G-LIKE 2.
 FT DOMAIN 2510 2736 LAMININ G-LIKE 3.
 FT DOMAIN 2737 2913 LAMININ G-LIKE 4.
 FT DOMAIN 2914 3075 LAMININ G-LIKE 5.
 FT DOMAIN 1706 1796 COILED COIL (POTENTIAL).
 FT DOMAIN 1968 1989 COILED COIL (POTENTIAL).
 FT DOMAIN 2088 2120 COILED COIL (POTENTIAL).
 FT SITE 2534 2536 CELL ATTACHMENT SITE.
 FT DISULFID 270 279 BY SIMILARITY.
 FT DISULFID 272 290 BY SIMILARITY.
 FT DISULFID 292 301 BY SIMILARITY.
 FT DISULFID 297 305 POTENTIAL.
 FT DISULFID 304 324 BY SIMILARITY.
 FT DISULFID 327 336 BY SIMILARITY.
 FT DISULFID 329 361 BY SIMILARITY.
 FT DISULFID 364 373 BY SIMILARITY.
 FT DISULFID 376 394 BY SIMILARITY.
 FT DISULFID 397 409 BY SIMILARITY.
 FT DISULFID 399 427 BY SIMILARITY.
 FT DISULFID 429 438 BY SIMILARITY.
 FT DISULFID 441 451 BY SIMILARITY.
 FT DISULFID 454 467 BY SIMILARITY.
 FT DISULFID 466 471 BY SIMILARITY.
 FT DISULFID 473 482 BY SIMILARITY.
 FT DISULFID 485 500 BY SIMILARITY.
 FT DISULFID 742 751 BY SIMILARITY.
 FT DISULFID 757 769 BY SIMILARITY.
 FT DISULFID 760 769 BY SIMILARITY.
 FT DISULFID 772 788 BY SIMILARITY.
 FT DISULFID 791 806 BY SIMILARITY.
 FT DISULFID 793 816 BY SIMILARITY.
 FT DISULFID 819 828 BY SIMILARITY.
 FT DISULFID 831 846 BY SIMILARITY.
 FT DISULFID 849 863 BY SIMILARITY.
 FT DISULFID 851 870 BY SIMILARITY.
 FT DISULFID 873 882 BY SIMILARITY.
 FT DISULFID 885 899 BY SIMILARITY.
 FT DISULFID 902 914 BY SIMILARITY.
 FT DISULFID 904 921 BY SIMILARITY.
 FT DISULFID 923 932 BY SIMILARITY.
 FT DISULFID 935 948 BY SIMILARITY.
 FT DISULFID 951 963 BY SIMILARITY.
 FT DISULFID 953 969 BY SIMILARITY.
 FT DISULFID 971 980 BY SIMILARITY.
 FT DISULFID 983 995 BY SIMILARITY.
 FT DISULFID 998 1007 BY SIMILARITY.
 FT DISULFID 1000 1014 BY SIMILARITY.

```

FT DISULFID 1016 1025 BY SIMILARITY.
FT FT 1028 1041 BY SIMILARITY.
FT DISULFID 1044 1056 BY SIMILARITY.
FT FT 1063 1063 BY SIMILARITY.
FT DISULFID 1065 1074 BY SIMILARITY.
FT FT 1087 1087 BY SIMILARITY.
FT DISULFID 1403 1412 BY SIMILARITY.
FT FT 1419 1419 BY SIMILARITY.
FT DISULFID 1420 1431 BY SIMILARITY.
FT FT 1449 1449 BY SIMILARITY.
FT DISULFID 1452 1466 BY SIMILARITY.
FT FT 1476 1476 BY SIMILARITY.
FT DISULFID 1479 1488 BY SIMILARITY.
FT FT 1506 1506 BY SIMILARITY.
FT DISULFID 1491 1521 BY SIMILARITY.
FT FT 1528 1528 BY SIMILARITY.
FT DISULFID 1511 1539 BY SIMILARITY.
FT FT 1530 1539 BY SIMILARITY.
FT DISULFID 1542 1556 INTERCHAIN (PROBABLE).
FT FT 1556 1556 INTERCHAIN (PROBABLE).
FT DISULFID 1560 1560 INTERCHAIN (PROBABLE).
FT FT 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 555 555 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT 665 665 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 763 763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT 926 926 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 952 952 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT 1045 1045 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1407 1407 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT 1579 1579 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1596 1596 N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 6.9%; Score 175.5; DB 1; Length 3075;
Best Local Similarity 21.2%; Pred. No. 6.5e-05;
Matches 110; Conservative 44; Mismatches 187; Indels 177; Gaps 30;

QY 18 IKSANCPVGTETAGQVDDLTGTPANCVCOKNFYNNAAAFVPGASTCTPCPKKDKAGA 77
DB 706 VECECPQGY-----TGTSCECLSGYRVDGILF---GGICQPCCECHGA--- 748
QY 78 QPNPPATANLVTCNVKCPAGTAIGAGTDYAAIITECVNCYRNFYENAPNFAGASTC 137
DB 749 -----AECNVH---GVCI---ACAHNTGVHCBOCLPGFGEFSGRTGDCQPC 791
QY 138 TACPVRVGGALTAGNAATIVACQNVACPTGTALDDG---VTTDYVR---SFTVCVKRLN 192
DB 792 -ACPL-----TIAS--NFESPT-CHLNDGDEVVCDWCAPGYSGAMCERCADG 834
QY 193 FYFNGNNGTFFNPGKSQCTPCPAIKPAKPAQATLGNDAITTAQCNVACPDGTISAAGVN 252
DB 835 YYGN-----PTVPGES-CVPCDC--SGNVDPSEAGHCDSTVTEC-LKCLGNTDGA--- 880
QY 253 NWVAQNTCTNAPNEYN--NAPN-----FNPGNSTCLPCPANKR-- 291
DB 881 -----HCERCADGFTYDVAATKNCRACECHVKGSHSACHVLETLGLCDCKPNVTGQCD 933
QY 292 -----XGAETAGGAATLAKQCNACPDGTATASGATNY-----VILQTECLNCAAN 338
DB 934 QCLHGYGLDSSGHG-----CRPCN--CSVAGSVSDGCTDEGQCHCVGAGKRCRCAHG 986
QY 339 FYFGNNGFAGSSRCKACPAKPVGATAGGTATLTAQCALECPAGT-----VLTDGT 392
DB 987 FY----AYQDGS--CTPCDCPHTQNTCDPTG-----ECV--CPPTQSGKCECEDGH 1032
QY 393 TSTYKQAASECVKCAANFYTKQTDWVAGIDTCTG-----C 428
DB 1033 WGYDEVCCQACNSLVGSTHRCDDVTGHCCKSKGGRACDQCSGLGYRDFPCVPCDC 1092
QY 429 NKKLTSGAEANLPES-----AKNI-----QCD 451
DB 1093 DLRGTSGDACNLQGLGCGVEETGACPCCKENVFQPCN 1130

RESULT 13

```

```

LMA5_MOUSE
ID LMA5_MOUSE STANDARD; PRT; 3635 AA.
AC Q61001;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE LAMININ ALPHA-5 CHAIN (FRAGMENT).
GN LAMA5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX MEDLINE=96081906; PubMed=7499364;
RA Miner J.H., Lewis R.M., Sanes J.R.;
RT "Molecular cloning of a novel laminin chain, alpha 5, and widespread
RT expression in adult mouse tissues.";
RL J. Biol. Chem. 270:28523-28526(1995).
RN [2]
REVISTONS.
RA Miner J.H., Lewis R.M., Sanes J.R.;
RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RL FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF
CC ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.
CC SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC MEMBRANES (MAJOR COMPONENT).
CC TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND
CC KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT
CC AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.
CC DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC SIMILARITY: ALPHA-5 CHAIN IS MORE RELATED TO DROSOPHILA ALPHA
CC CHAIN THAN TO MAMMALIAN ALPHA CHAINS 1-4.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U37501; AAC53430.1; .
HSSP; P02468; TITLE.
MGD; MGI:105382; LAMA5.
InterPro; IPR000361; EGF-like.
InterPro; IPR001886; LamNT.
InterPro; IPR000034; Laminin_B.
InterPro; IPR002049; Laminin_EGF.
InterPro; IPR001791; Laminin_G.
Pfam; PF00052; laminin_B_1.
Pfam; PF00053; laminin_EGF; 19.
Pfam; PF00054; laminin_G; 2.
Pfam; PF00055; laminin_Nterm; 1.
ProDom; PD002082; LamNT; 1.
ProDom; PD003031; Laminin_B; 1.
SMART; SM00180; EGF_Lam; 17.
SMART; SM00001; EGF_like; 2.

```



```
DR HSP: Q63450; 1A06.
DR WormPep; D1044.3; CE01206.
DR InterPro; IPR002899; EB.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_kin_actsite.
DR Pfam; PF01683; EB; 12.
DR Pfam; PF00069; pkinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00289; WRI; 12.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS01186; EGF_2; 2.
KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;
KW ATP-binding; EGF-like domain. PROTEIN KINASE.
FT DOMAIN 431 703
FT NP_BIND 437 445 ATP (BY SIMILARITY).
FT BINDING 461 461 ATP (BY SIMILARITY).
FT ACT_SITE 569 569 BY SIMILARITY.
FT ACT_SITE 569 569 BY SIMILARITY.
SQ SEQUENCE 1895 AA; 208393 MW; F23C9F7881353AD6 CRC64;

Query Match 6.7%; Score 171; DB 1; Length 1895;
Best Local Similarity 20.3%; Pred. No. 8.5e-05;
Matches 111; Conservative 53; Mismatches 220; Indels 162; Gaps 25;

QY 3 NNILVILII---SLFINQ--IKSANCVPGETNTAGQVDDLQ--TPANCVNCOKNFYNN 55
DB 1190 NNQCVLLSIVGETCIANQCCVGGAMCSGTCQCTNGATAMYGVCISSSSSCSNSNQVSIN 1249

QY 56 AAAF--VPGASTCTPCQKDGAGQNPAPTANLV--TQCNVKCPAG-TAAGGATDYAA 110
DB 1250 GMCYNTVQVGGSCFSQ-----CLNNAVCINNICTVSTFCSVCSSTNQVCISNQCYNVS 1304

QY 111 IITEVCNCRINFYNENAPFNAGASTCTACPVNRVGGALTAGNAATIVAQCNVACPTGTA 170
DB 1305 IGSQCV-----GSQQCLS-----NSQCISISICQ--CPQGTQ 1333

QY 171 LDDGYTIDYRSFTCEVKRLNFYNGNGNGTFFNPGK-----SOCTPCPAIK-PANVAQA 225
DB 1334 QSNGVCT-----GNNNNNQCPNOVLINNOCYNTVSGFQCQFPQ 1375

QY 226 TLGNDATITACQNVACPDGTISAAGV-----NNWVAQTECTNCAPNFYNNNAPN 275
DB 1376 CLGNSQCNSMCQ--CPTGSTNVNGYCOGSGNGCNSQVLIINNOCYNTVSGFCQCPAQ 1433

QY 276 FNPGNSTCL----PCPANKDYGAETAGGA-----ATLAKOCNIA-- 311
DB 1434 QCLGNSQCLNSICQCPSGSSNVNGYCOGSGNGCNSNOVYNNOCYNTVPIGSQCQITQQ 1493

QY 312 -----CPDGT-----IASGATNVVILQTECLNCAANFYFDG 343
DB 1494 CLGNSQCMNSFCQCPSGPTNVNFCTSSSSSNCLSAGCTVOLDSSNOPINCLYSTCPNN 1553

QY 344 NNFOAGSSR-----CKACPANKVQGAATAGTATLIAQCALECPAGTVLTDGTTSTYKQ 398
DB 1554 SFCQYSSSQRYVCCRKCCTNS-SPOVSASQGVVTCFTNS--QCASGYICSNGAC----- 1605

QY 399 AASECVKCAANFYTTKQTDWVAG-IDTCTSCNKKLTSGAEANLPESAKKNIQCD----- 451
DB 1606 ----CPNTNSNTCSTGTGTCFTGQISVSGQCFCFNSVIGDRQQRSEQCLGGSCQNNLCQC 1661

QY 452 ----FAN 454
DB 1662 PNGFAN 1667
```